

## #1. DOWNLOAD THE DATASET

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

## #2. LOADING THE DATASET

```
df = pd.read_csv('/content/House Price India.csv')
df.head()
```

	<b>id</b>	<b>Date</b>	<b>number_of_bedrooms</b>	<b>number_of_bathrooms</b>	<b>living_area</b>	<b>lot_area</b>	
<b>0</b>	6762810145	42491		5	2.50	3650	9050
<b>1</b>	6762810635	42491		4	2.50	2920	4000
<b>2</b>	6762810998	42491		5	2.75	2910	9480
<b>3</b>	6762812605	42491		4	2.50	3310	42998
<b>4</b>	6762812919	42491		3	2.00	2710	4500

5 rows × 23 columns

```
df.shape
```

```
(14620, 23)
```

```
# 3 . PERFORMING THE UNI-VARIANT , BI-VARIANT , MULTI VARIANT ANALYSIS OVER THE HOUSE PRICE DATASET
sns.distplot(df['lot_area'])
```

```
<ipython-input-8-9662b0fcfa36b>:2: UserWarning:
```

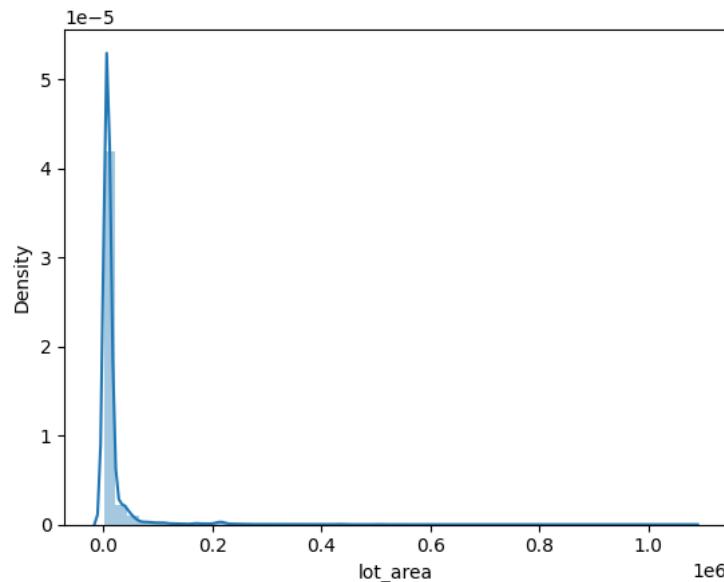
```
`distplot` is a deprecated function and will be removed in seaborn v0.14.0.
```

```
Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
```

```
For a guide to updating your code to use the new functions, please see
```

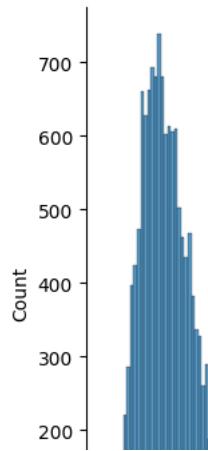
```
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

```
sns.distplot(df['lot_area'])
<Axes: xlabel='lot_area', ylabel='Density'>
```



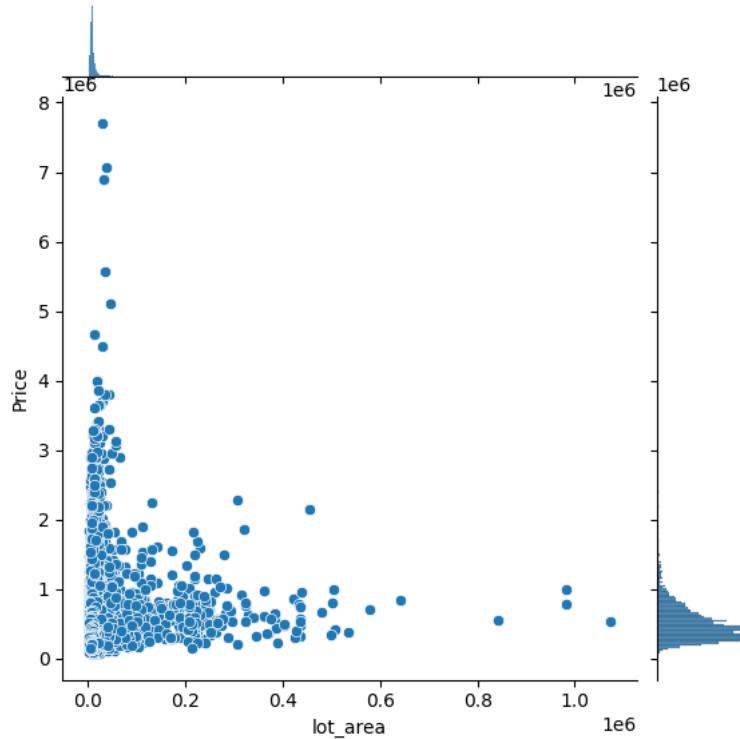
```
sns.distplot(df['living_area'])
```

```
<seaborn.axisgrid.FacetGrid at 0x7a7511622200>
```



```
sns.jointplot(x='lot_area',y='Price',data=df)
```

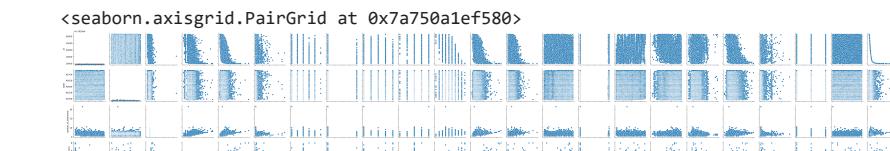
```
<seaborn.axisgrid.JointGrid at 0x7a750b7f78b0>
```



```
df.corr()
```

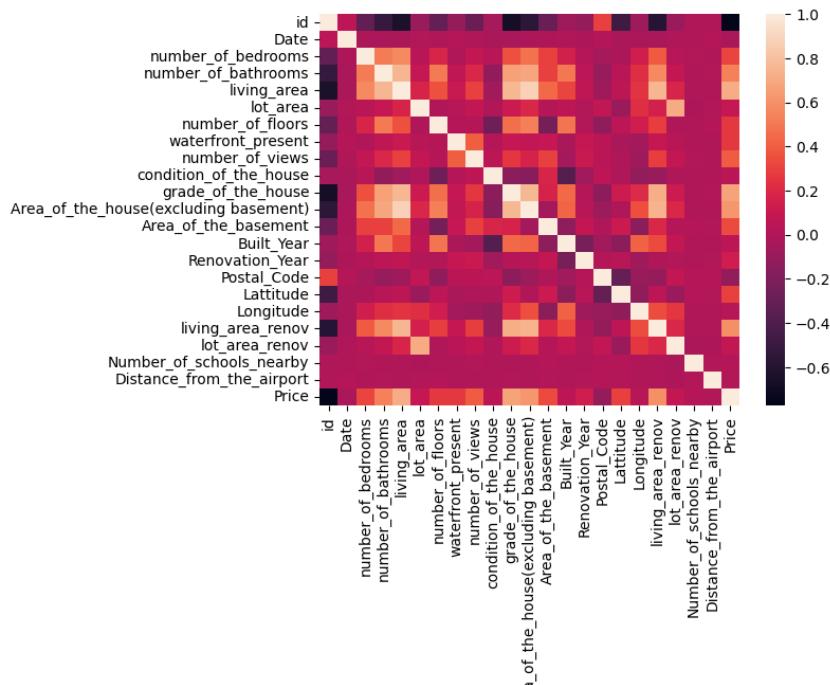
	<b>id</b>	<b>Date</b>	<b>number_of_bedrooms</b>	<b>number_of_bathrooms</b>
<b>id</b>	1.000000	0.045966	-0.329034	-0.1
<b>Date</b>	0.045966	1.000000	-0.015663	-0.0
<b>number_of_bedrooms</b>	-0.329034	-0.015663	1.000000	0.1
<b>number_of_bathrooms</b>	-0.516909	-0.026485	0.509784	1.0
<b>living_area</b>	-0.648127	-0.021958	0.570526	0.1
<b>lot_area</b>	-0.100269	0.004392	0.034416	0.0
<b>number_of_floors</b>	-0.312305	-0.010335	0.177294	0.1
<b>waterfront_present</b>	-0.112937	0.012006	-0.006257	0.0
<b>number_of_views</b>	-0.293004	-0.004782	0.078665	0.1
<b>condition_of_the_house</b>	-0.045061	-0.027402	0.026597	-0.1

```
sns.pairplot(df)
```



```
sns.heatmap(df.corr())
```

<Axes: >



#4 descriptive statistics

```
df.describe()
```

	<b>id</b>	<b>Date</b>	<b>number_of_bedrooms</b>	<b>number_of_bathrooms</b>	<b>living</b>
<b>count</b>	1.462000e+04	14620.000000	14620.000000	14620.000000	14620.0
<b>mean</b>	6.762821e+09	42604.538646	3.379343	2.129583	2098.2
<b>std</b>	6.237575e+03	67.347991	0.938719	0.769934	928.2
<b>min</b>	6.762810e+09	42491.000000	1.000000	0.500000	370.0
<b>25%</b>	6.762815e+09	42546.000000	3.000000	1.750000	1440.0
<b>50%</b>	6.762821e+09	42600.000000	3.000000	2.250000	1930.0
<b>75%</b>	6.762826e+09	42662.000000	4.000000	2.500000	2570.0
<b>max</b>	6.762832e+09	42734.000000	33.000000	8.000000	13540.0

8 rows × 23 columns

◀ ▶

#5. missing values

```
df.isnull().any()
```

<b>id</b>	False
<b>Date</b>	False
<b>number_of_bedrooms</b>	False
<b>number_of_bathrooms</b>	False
<b>living_area</b>	False
<b>lot_area</b>	False
<b>number_of_floors</b>	False
<b>waterfront_present</b>	False

```
number_of_views      False
condition_of_the_house  False
grade_of_the_house    False
Area_of_the_house(excluding basement)  False
Area_of_the_basement  False
Built_Year            False
Renovation_Year       False
Postal_Code           False
Latitude              False
Longitude             False
living_area_renov     False
lot_area_renov        False
Number_of_schools_nearby  False
Distance_from_the_airport  False
Price                 False
dtype: bool
```

Colab paid products - [Cancel contracts here](#)

✓ 1s completed at 9:46 PM

